

Priya Ranganathan,
C. S. Pramesh¹,
Marc Buyse²

Department of Anaesthesiology, Tata Memorial Centre,
¹Department of Surgical Oncology, Division of Thoracic Surgery, Tata Memorial Centre, Mumbai, Maharashtra, India, ²International Drug Development Institute, Hasselt University, Diepenbeek, Belgium

Address for correspondence:

Dr. Priya Ranganathan,
Department of Anaesthesiology,
Tata Memorial Centre, Dr. Ernest Borges Road, Parel, Mumbai - 400 012, Maharashtra, India.
E-mail: drpriyaranganathan@gmail.com

Common pitfalls in statistical analysis: “P” values, statistical significance and confidence intervals

Abstract

In the second part of a series on pitfalls in statistical analysis, we look at various ways in which a statistically significant study result can be expressed. We debunk some of the myths regarding the 'P' value, explain the importance of 'confidence intervals' and clarify the importance of including both values in a paper

Key words: Biostatistics, bias (epidemiology), confidence interval

The objective of a superiority trial is to demonstrate increased efficacy of one treatment over another. The statistical implications of the results of such trials can be reported in two ways. The “P” value provides the probability that the observed result occurred by chance. The “P” value needs to be interpreted in the context of the “alpha” or type 1 error, which is decided before the study is commenced. The type 1 error is the probability of finding a difference between treatments by chance, when a difference does not actually exist. It is conventionally set at 5%, which means that if the study finds a difference between treatments, we can be 95% sure that this is a true difference and not a chance finding. For a type 1 error of 5%, the corresponding “P” value should be <0.05 to be considered statistically significant. The smaller the “P”

value, the less likely that the findings of the study are due to chance. A common misconception is that smaller P values imply that the treatment is more effective than if the P value were higher. This is only true for a given sample size. Hence, it is inappropriate to compare the P values of trials of different sizes. For any given sample size, a P value of 0.05 merely implies that the possibility of the findings of the study being due to chance (i.e. no real difference between the two groups) is 5% whereas a P value of 0.01 implies that the possibility of the study findings being due to chance is 1%. How effective the treatment is should not be assessed through the P value but rather through the difference in means (or proportions) between the two treatments under comparison. Similarly, the P value does not by itself indicate which of the treatments compared is superior; it merely implies that one treatment is superior to the other. From the above, it is clear that P values, by themselves, do not indicate either the magnitude or the direction of the difference.

The difference in means, in contrast, does not provide information on the uncertainty of the observed result. A better way of expressing study results is to provide

Access this article online

Quick Response Code:



Website:

www.picronline.org

DOI:

10.4103/2229-3485.154016

the confidence interval (CI) in addition to the observed difference. Study results are derived from a single sample, which is considered to be representative of the population. CIs give the range of possible values for the measured variable that might be found if the study was repeated on multiple samples of the same size drawn from the population. They are a measure of the precision of the study results when extrapolating them to the population from which the study sample was drawn. In other words, CIs give a range where the true value (in the population) is likely to lie, with a certain probability. While comparing two groups, CIs also give the direction and strength of the probable effect.^[1] For example, if the mean and 95% CIs of the systolic blood pressure in a study sample is 130 mmHg (95% CI, 115–142), we are 95% sure that the true mean of systolic blood pressure in the population lies between 115 mmHg and 142 mmHg). When comparing two groups, if the CI of the difference between the groups does not include the value of “no effect,” then this automatically implies statistical significance.^[2] While comparing two group means, the value of “no effect” is zero and hence, for a type 1 error of 5%, the 95% CIs for the difference between the means should not include the value “zero.” This implies that we can be 95% sure that the difference between the means is not zero, or that there is a true difference between the two means. While comparing two proportions (or ratios), the value of “no effect” is one. And hence, CIs for a difference between proportions (odds ratio, the relative risk) not including the value “one” is considered statistically significant.

Corneli *et al.* conducted a randomized trial to compare the effect of dexamethasone versus placebo in children with bronchiolitis.^[3] The hospital admission rate was 39.7% for children assigned to dexamethasone, compared with 41.0% for those assigned to placebo (absolute difference,

–1.3%; $P = 0.74$). This P value suggests that there is a 74% possibility that the difference in hospital admission occurred only by chance and is considered statistically not significant. The 95% CI for this difference was –9.2% to +6.5%. This means that we are 95% sure that in the population, dexamethasone could either reduce the admission rate by as much as 9.2% (benefit) or actually increase the admission rate by as much as 6.5% (harm). These CIs also include the value “zero” which means a possibility that there is no difference between the two treatments. Therefore, we cannot be confident about the benefit of dexamethasone over placebo.

To summarize, CIs provide a range of values where we can be reasonably confident the truth lies, the direction as well as the magnitude of the difference between two groups as well as statistical significance; P values on the other hand, quantify the probability of the study findings being due to chance (the strength of the evidence), but provide no direct measure of the magnitude or direction of the effect. Hence, while CIs provide more information than P values, the two are complementary, and authors should report both of these in their papers.

REFERENCES

1. du Prel JB, Hommel G, Röhrig B, Blettner M. Confidence interval or P value?: Part 4 of a series on evaluation of scientific publications. *Dtsch Arztebl Int* 2009;106:335-9.
2. Gupta SK. The relevance of confidence interval and P value in inferential statistics. *Indian J Pharmacol* 2012;44:143-4.
3. Corneli HM, Zorc JJ, Mahajan P, Shaw KN, Holubkov R, Reeves SD, *et al.* A multicenter, randomized, controlled trial of dexamethasone for bronchiolitis. *N Engl J Med* 2007;357:331-9.

How to cite this article: Ranganathan P, Pramesh CS, Buyse M. Common pitfalls in statistical analysis: “ P ” values, statistical significance and confidence intervals. *Perspect Clin Res* 2015;6:116-7.

Source of Support: Nil. **Conflict of Interest:** None declared.